Stochastic Processes and their Applications 84 (1999) 343–355

stochastic processes and their applications

www.elsevier.com/locate/spa

Sequential point estimation of parameters in a threshold AR(1) model

Sangyeol Lee[a,*], T.N. Sriram[b]

[a] *Department of Statistics, Seoul National University, Seoul, 151-742, South Korea*

[b] *Department of Statistics, University of Georgia, Athens, GA 30602, USA*

Received 16 November 1998; received in revised form 2 June 1999

Abstract

We show that if an appropriate stopping rule is used to determine the sample size when estimating the parameters in a stationary and ergodic threshold AR(1) model, then the sequential least-squares estimator is asymptotically risk efficient. The stopping rule is also shown to be asymptotically efficient. Furthermore, non-linear renewal theory is used to obtain the limit distribution of appropriately normalized stopping rule and a second-order expansion for the expected sample size. A central result here is the rate of decay of lower-tail probability of average of stationary, geometrically β-mixing sequences.

MSC: Primary 62L12; 62M10; Secondary 62F10; 62F12

Keywords: TAR models; Ergodicity; Asymptotic risk efficiency; Asymptotic efficiency; Uniform integrability; Geometrically β-mixing; Stopping rule

1. Introduction

The idea (due to Robbins, 1959) of using stopping rules to determine precisely the right sample size needed to construct a point estimator has been used extensively in the literature. For the independent and identically distributed (i.i.d.) setup, theoretical performance of sequential point estimators has been studied in detail and these are lucidly described in Woodroofe (1982), Martinsek (1983), and Ghosh et al. (1997). In the last decade, this intuitive idea has been extended to models with dependent data, such as linear time series and branching processes. See, for instance, Sriram (1987,1988) and Sriram et al. (1991) for sequential estimation problems arising in autoregressive (AR) model of order 1 and branching processes, respectively, and Fakhre-Zakeri and Lee (1992,1993) and Lee (1994,1996) for extension of Sriram's results to AR(p) models and linear processes.

It is now natural to ask whether sequential methods can be extended to estimation problems arising in nonlinear time series models. Recently, Sriram (1998) proposed a

* Corresponding author.

E-mail address: sylee@stats.snu.ac.kr (S. Lee)

0304-4149/99/$ - see front matter
PII: S0304-4149(99)00060-5

stopping rule to construct a sequential fixed-size confidence ellipsoid for the parameters in threshold autoregressive (TAR) models. The purpose of this paper, however, is to construct a point estimator based on a stopping rule for the estimation of parameters in TAR models. Asymptotic properties of these sequential procedures are then studied.

Sequential point estimation problems generally require a study of growth rate of mean-squared error (MSE) of estimators. Since the estimators involved in time-series setup are usually of ratio-type, in order to obtain the growth rate of MSE, one needs to the study the rate of convergence of lower-tail probability of the denominator variables. In his work, Lee (1994) used the linearity of AR(p) time series (among other things) crucially to obtain such lower-tail probability rates. Unfortunately, since TAR models are nonlinear, we cannot duplicate the techniques from Lee (1994). We adopt alternative methods of obtaining such lower-tail probability rates (see Proposition 2.1 and Lemma 2.2 below) and thus obtain a rate for MSE. We believe that the techniques adopted here will be useful more generally.

Threshold models, introduced by Tong (1978a,b), are generally agreed to be useful in modeling discrete time series that exhibit piece-wise linearity. In fact, Tong and Lim (1980) provide many examples where TAR models not only provide a better fit than linear models but also exhibit strictly nonlinear behavior (e.g., limit cycles, jump resonance, harmonic distortion, etc.) which linear models cannot duplicate. For a comprehensive study of threshold models and other nonlinear models, see Tong (1983,1990).

A TAR (1) process $\{X_i\}$ is defined by

$$X_i = \theta_1 X_{i-1}^{+} + \theta_2 X_{i-1}^{-} + \varepsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ are real parameters not necessarily equal, $\{\varepsilon_i\}$ is a sequence of i.i.d. random variables (r.v.'s), and $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$ for a real number x. Throughout it is assumed that $E\varepsilon_1 = 0 < E\varepsilon_1^2 = \sigma^2 < \infty$ where σ^2 is an unknown constant and the distribution of ε_1 is unspecified.

It has been shown in Petruccelli and Woolford (1984) that the process $\{X_i; \geqslant 0\}$ defined in (1.1) is ergodic if and only if

$$\boldsymbol{\theta} \in \Theta = \left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} : \theta_1 < 1, \theta_2 < 1, \theta_1\theta_2 < 1 \right\}. \tag{1.2}$$

This implies the existence of an invariant probability distribution for $\{X_i\}$. We shall assume that the initial random variable X_0 has its distribution $\pi(\cdot)$ the invariant probability distribution of the Markov chain $\{X_i\}$ so that the process $\{X_i\}$ is strictly stationary. Also, we note from Chan et al. (1985) that $E|\varepsilon_1|^k < \infty$ for some integer $k \geqslant 1$ implies that $E|X_0|^k < \infty$ for each $\boldsymbol{\theta} \in \Theta$.

We are interested in the problem of point estimation of the parameters θ_1 and θ_2 in (1.1). Suppose we estimate the parameters θ_1 and θ_2 in (1.1) by their least-squares estimators

$$\hat{\theta}_{1,n} = \sum_{i=1}^{n} X_i X_{i-1}^{+} \Big/ \sum_{i=1}^{n} X_{i-1}^{+2} \tag{1.3}$$

and

$$\hat{\theta}_{2,n} = \sum_{i=1}^{n} X_i X_{i-1}^{-} \Big/ \sum_{i=1}^{n} X_{i-1}^{-2} \tag{1.4}$$

subject to the loss function

$$L_n = An^{-1}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' \Gamma_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + n, \tag{1.5}$$

where $\hat{\boldsymbol{\theta}}_n = \binom{\hat{\theta}_{1,n}}{\hat{\theta}_{2,n}}$ and $\Gamma_n = \mathrm{diag}(\sum_{i=1}^{n} X_{i-1}^{+2}, \sum_{i=1}^{n} X_{i-1}^{-2})$ is a diagonal matrix. In (1.5), $A(>0)$ reflects the importance of quadratic error relative to sampling cost which is assumed to be one unit per observation. Our objective here is to minimize the risk in estimation by choosing an appropriate sample size.

From Theorem 3.2 of Petruccelli and Woolford (1984) it follows that

$$\sigma^{-2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' \Gamma_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{D}} \chi_2^2 \quad \text{as } n \to \infty, \tag{1.6}$$

where χ_2^2 is a chi-square random variable with two degrees of freedom. Suppose for the moment that the sequence

$$\{Q_n = (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' \Gamma_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}); n \geqslant 2\} \text{ is uniformly integrable (u.i.).} \tag{1.7}$$

The result in (1.7), incidentally, is established in Proposition 2.1 below under certain moment conditions. Now, (1.6) together with (1.7) yields

$$R_n = EL_n = 2n^{-1}A\sigma^2 + n + \mathrm{o}(n^{-1}) \quad \text{as } n \to \infty. \tag{1.8}$$

It can then be shown that the risk R_n is approximately minimized by

$$n_0(A) \approx (2A)^{1/2}\sigma \tag{1.9}$$

with the corresponding minimum risk

$$R_{n_0(A)} \approx 2(2A)^{1/2}\sigma. \tag{1.10}$$

However, when σ^2 is unknown, $n_0(A)$ cannot be used in practice and there is no fixed sample size that will achieve the minimum risk (1.10).

To overcome this, we replace the unknown σ^2 by its least-squares estimator $\hat{\sigma}_n^2$ defined by

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} (X_i - \hat{\theta}_{1,n} X_{i-1}^{+} - \hat{\theta}_{2,n} X_{i-1}^{-})^2. \tag{1.11}$$

Now, we mimick the nature of $n_0(A)$ and define a stopping rule T_A by

$$T_A = \inf\{n \geqslant n_A \colon n \geqslant (2A)^{1/2} \hat{\sigma}_n\}, \tag{1.12}$$

where n_A is an initial sample size possibly depending on A. The aim here is to assess the performance of the stopping rule T_A, ET_A, and the risk of the sequential procedure $R_A = EL_{T_A}$ as the penalty $A \to \infty$. The following results are established.

Theorem 1. *Suppose that* $E|\varepsilon_1|^{4p} < \infty$ *for* $p > 2$ *and* $\boldsymbol{\theta} \in \Theta$ *defined in* (1.2). *Let the initial sample size* n_A *be such that* $A^{1/2(1+\eta)} \leqslant n_A = \mathrm{o}(A^{1/2})$ *with* $\eta \in (0, (p-2)/2)$. *Then, as* $A \to \infty$

$$T_A / n_0(A) \to 1 \text{ a.s.}, \tag{1.13}$$

$$E|T_A/n_0(A) - 1| \to 0 \tag{1.14}$$

$$R_A/R_{n_0(A)} \to 1. \tag{1.15}$$

Theorem 2. *If* $E|\varepsilon_1|^4 < \infty$ *and in addition* $n_A = o(A^{1/2})$, *then*

$$[T_A - (2A)^{1/2}\sigma]/\sqrt{(2A)^{1/2}\sigma} \xrightarrow{D} N(0, \alpha^2), \tag{1.16}$$

where $\alpha^2 = (1/4)Var(\varepsilon_1^2/\sigma^2)$. *Furthermore, under the conditions of Theorem* 1, *if the distribution of* ε_1 *is nonarithmetic then*

$$ET_A - (2A)^{1/2}\sigma = c_1 + o(1) \quad \text{as } A \to \infty, \tag{1.17}$$

where the constant c_1 *is given in* (3.16) *below.*

The result in (1.14) says that, as the penalty for estimation error tends to infinity, the stopping rule defined in (1.12) is asymptotically efficient while (1.15) says that the associated sequential procedure is asymptotically risk efficient. Result in (1.16) describes the limiting distribution of T_A while (1.17) explicitly evaluates the second-order behavior of ET_A with $n_0(A)$ as its leading term. Proof of Theorem 1 along with some rate of convergence results which are of independent interest are given in Section 2. Theorem 2 is proved in Section 3 using nonlinear renewal theory developed by Lai and Siegmund (1977,1979) and Hagwood and Woodroofe (1982).

2. Basic convergence results

A central result of the paper is Proposition 2.1 where a crucial rate of convergence for the lower-tail probability of the sequences $\{n^{-1}\sum_{i=1}^n X_{i-1}^{\pm 2}; n \geqslant 1\}$ is obtained. This is then used to establish the uniform integrability of $\{Q_n; n \geqslant 1\}$ defined in (1.7). The former result is derived as a consequence of a general result proved in Lemma 2.2 which obtains a rate of convergence for the lower-tail probability of average of bounded, geometrically β-mixing, stationary r.v.'s. In addition, we also obtain rate of convergence of tail behavior of T_A. All these results are then used to prove Theorem 1. First we note the following: For $\hat{\sigma}_n^2$ defined in (1.11)

$$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \varepsilon_i^2 - n^{-1}Q_n, \tag{2.1}$$

where Q_n is as defined in (1.7). Consequently,

$$\max\{Q_n, n\hat{\sigma}_n^2\} \leqslant \sum_{i=1}^n \varepsilon_i^2. \tag{2.2}$$

Throughout this paper $\mathscr{F}_n = \sigma\{X_0, \varepsilon_1, \ldots, \varepsilon_n\}$ and $||\cdot||_p$ denotes the L_p-norm. The following lemma gives a rate of convergence of the L_p norm of numerator r.v.'s in (1.3) and (1.4).

Lemma 2.1. *If* $E|\varepsilon_1|^p < \infty$ *for* $p \geqslant 2$, *then for* $0 < u \leqslant v < \infty$, *the following results hold*:

$$\left\| \max_{[un] \leqslant l \leqslant [vn]} \left| l^{-1/2} \sum_{i=1}^{l} X_{i-1}^{\pm} \varepsilon_i \right| \right\|_p = \mathrm{O}(1) \quad \text{as } n \to \infty, \tag{2.3}$$

where $X_i^{\pm}$ *denotes either* X_i^{+} *or* X_i^{-}, *and the sequence*

$$\left\{ n^{-1/2} \sum_{i=1}^{n} X_{i-1}^{\pm} \varepsilon_i ; n \geqslant 1 \right\} \tag{2.4}$$

is uniformly continuous in probability (*u.c.i.p.*) *and stochastically bounded.*

Proof. Under the assumptions in (1.1) it is easily checked that $\{\sum_{i=1}^{n} X_{i-1}^{\pm} \varepsilon_i, n \geqslant 1\}$ is a martingale sequence with respect to $\mathscr{F}_n$ defined above. Now, as in the proof of Lemma 1 of Lee (1994), use the Doob's maximal inequality, the Burkhölder's inequality (see, Chow and Teicher, 1978, Theorem 11.2.1), the moment inequality (with $p \geqslant 2$), and the stationarity of $\{X_i\}$ to get the result in (2.3).

The stochastic boundedness of $\{n^{-1/2} \sum_{i=1}^{n} X_{i-1}^{\pm} \varepsilon_i ; n \geqslant 1\}$ follows from (2.3). The u.c.i.p. of $\{n^{-1/2} \sum_{i=1}^{n} X_{i-1}^{\pm} \varepsilon_i ; n \geqslant 1\}$ can be proved using arguments in Sriram (1988) or Lee (1994). □

Proposition 2.1. *For every* $0 < \delta < E(X_0^{\pm})^2$, *there is a* $\lambda \in (0,1)$ *such that*

$$P\left\{ n^{-1} \sum_{i=1}^{n} X_{i-1}^{\pm 2} < \delta \right\} = \mathrm{O}(n^{5/2} \lambda^{\sqrt{n}}) \quad \text{as } n \to \infty. \tag{2.5}$$

Moreover, if in addition, $E|\varepsilon_1|^{4s} < \infty$ *for some* $s \geqslant 1$, *then for all* $0 < u \leqslant v < \infty$

$$\sup_{n \geqslant 1} E \max_{[un] \leqslant l \leqslant [vn]} Q_l^s < \infty \tag{2.6}$$

and, consequently, for all $q < s$

$$\{Q_n^q ; n \geqslant 2\} \text{ is uniformly integrable.} \tag{2.7}$$

In particular, $EQ_n \to 2\sigma^2$ *as* $n \to \infty$, *and hence* (1.8) *holds.*

The following lemma concerns the rate of decay of lower-tail probability of average of bounded, geometrically β-mixing, stationary r.v.'s. This lemma will be used to prove Proposition 2.1. For a comprehensive study of mixing theory see Doukhan (1994).

Lemma 2.2. *Let* $\{Y_i ; i \geqslant 0\}$ *be a stationary process and* $\mathscr{G}_j = \sigma(Y_0, \ldots, Y_j)$, $j \geqslant 0$, *be a nondecreasing sequence of* σ*-fields. Assume further that* $0 \leqslant Y_i \leqslant 1$ *for all* $i \geqslant 0$, *and that the process* $\{Y_i\}$ *is geometrically* β*-mixing, that is,*

$$\begin{aligned} \beta_n &= E \left\| \sup_{k \geqslant 0} \sup_{V \in \mathscr{G}_{n+k}^*} |P(V | \mathscr{G}_k) - P(V)| \right\|_\infty \\ &= \mathrm{O}(\rho^n) \quad \text{as } n \to \infty, \end{aligned} \tag{2.8}$$

where $\mathscr{G}_j^* = \sigma(Y_j, Y_{j+1}, \ldots)$, *for a random variable* Y, $||Y||_\infty = \text{ess sup } |Y|$ *and* $\rho \in (0,1)$ *is some real number. Then for every* $a \in (0, EY_1)$, *there is a* $\lambda \in (0,1)$ *such that*

$$P\left(\sum_{i=1}^{n} Y_i < na\right) = \mathrm{O}(n^{5/2}\lambda^{\sqrt{n}}) \quad \text{as } n \to \infty. \tag{2.9}$$

Proof. Define a sequence $\{Y_{n,i}\}$ of r.v.'s by

$$Y_{n,i} = \sum_{j=1}^{n^2} [(j-1)/n^2] I_{\{(j-1)/n^2 \leqslant Y_i \leqslant j/n^2\}} \tag{2.10}$$

for $i \geqslant 1$. Let b be a number such that $a < b < EY_1$ and N be a positive integer such that $a < b < EY_{n,1} \leqslant EY_1$ for all $n \geqslant N$. Let $m = m(n) = [\sqrt{n}]$. It is then possible to find $l (\geqslant 1)$ and $r = 0, \ldots, m-1$ such that $n = lm + r$. For simplicity, take $l = 2k$ for $k \geqslant 1$ and $r = 0$, that is, $n = 2km$. Now, define r.v.'s $\{Z_i; 1 \leqslant i \leqslant n\}$ and $\{Z_i', 1 \leqslant i \leqslant n\}$ such that

$$\begin{aligned}
Z_1 &= \sum_{i=1}^{m} Y_{n,i}, \quad Z_1' = \sum_{i=m+1}^{2m} Y_{n,i} \\
Z_2 &= \sum_{i=2m+1}^{3m} Y_{n,i}, \quad Z_2' = \sum_{i=3m+1}^{4m} Y_{n,i} \\
&\vdots \qquad\qquad\qquad \vdots \\
Z_k &= \sum_{i=2(k-1)m+1}^{(2k-1)m} Y_{n,i}, \quad Z_k' = \sum_{i=(2k-1)m+1}^{2km} Y_{n,i}.
\end{aligned}$$

Then, since $|Y_i - Y_{n,i}| \leqslant 1/n^2$, for large n we can choose $a' \in (a,b)$ such that

$$\begin{aligned}
P\left\{\sum_{i=1}^{n} Y_i < na\right\} &\leqslant P\left\{\sum_{i=1}^{n} Y_{n,i} < na'\right\} \\
&\leqslant P\left\{\sum_{i=1}^{k} m^{-1} Z_i < ka'\right\} + P\left\{\sum_{i=1}^{k} m^{-1} Z_i' < ka'\right\} \\
&= I + II.
\end{aligned} \tag{2.11}$$

We will now show that $I = \mathrm{O}(n^{5/2}\lambda^{\sqrt{n}})$ as $n \to \infty$. To this end, let $\tilde{\mathscr{G}}_1 = \mathscr{G}_m, \ldots, \tilde{\mathscr{G}}_k = \mathscr{G}_{(2k-1)m}$. Then, for each j, $m^{-1} Z_j$ is $\tilde{\mathscr{G}}_j$ measurable and $0 \leqslant m^{-1} Z_j \leqslant 1$. Furthermore, for each j,

$$\begin{aligned}
&E|E(m^{-1} Z_j | \tilde{\mathscr{G}}_{j-1}) - E(m^{-1} Z_j)| \\
&\quad \leqslant m^{-1} \sum_{v=1}^{m} E|E(Y_{n,2(j-1)m+v} | \mathscr{G}_{[2(j-1)-1]m}) - EY_{n,2(j-1)m+v}| \\
&\quad = \mathrm{O}(n^2 \rho^m)
\end{aligned} \tag{2.12}$$

by (2.10) and (2.8). Since $EY_{n,1} \to EY_1$ as $n \to \infty$ and $EY_1 > b$, for large n it is possible to choose $\delta_0 > 0$ such that $EY_{n,1} - \delta_0 > b$. For this δ_0, define the set

$$S = \{|E(m^{-1}Z_j|\tilde{\mathscr{G}}_{j-1}) - E(m^{-1}Z_j)| < \delta_0 \quad \text{for } j = 1, \ldots, k\}.$$

Note that, on S

$$\sum_{j=1}^{k} E(m^{-1}Z_j|\tilde{\mathscr{G}}_{j-1}) = \sum_{j=1}^{k} E(m^{-1}Z_j) + \sum_{j=1}^{k} [E(m^{-1}Z_j|\tilde{\mathscr{G}}_{j-1}) - E(m^{-1}Z_j)]$$

$$\geqslant k(EY_{n,1} - \delta_0) > kb.$$

Now apply a Theorem of Freedman (1973) (see (4a) and Note) to get

$$P\left\{\sum_{j=1}^{k} m^{-1}Z_j < ka', S\right\}$$

$$\leqslant P\left\{\sum_{j=1}^{k} m^{-1}Z_j < ka', \sum_{j=1}^{k} E(m^{-1}Z_j|\tilde{\mathscr{G}}_{j-1}) > kb\right\}$$

$$\leqslant \mathrm{e}^{-k(b-a')^2/2b} = \mathrm{O}(\mathrm{e}^{-d\sqrt{n}}) \tag{2.13}$$

for some $d > 0$ where we used the fact that $k = n/(2m)$ and $m = [\sqrt{n}]$. On the other hand, by the Markov inequality and (2.12)

$$P(S^c) \leqslant \delta_0^{-1} \sum_{j=1}^{k} \mathrm{O}(n^2 \rho^m)$$

$$= \mathrm{O}(n^{5/2}\lambda^{\sqrt{n}}), \tag{2.14}$$

since $k = \mathrm{O}(\sqrt{n})$ and λ is some number in (0,1). Applying (2.13) and (2.14) to I in (2.11) we get that $I = \mathrm{O}(n^{5/2}\lambda^{\sqrt{n}})$. Similar arguments as above yield $II = \mathrm{O}(n^{5/2}\lambda^{\sqrt{n}})$. Hence the lemma *is established.* □

Proof of Proposition 2.1. Let δ be as in the proposition. Define $\tilde{X}_i^{+2} = X_i^{+2} I_{\{X_i^{+2} \leqslant K\}}$ for each $i \geqslant 0$ and $K \geqslant 1$. Choose K large enough and fix it so that $\delta_1 = E\tilde{X}_0^{+2} > \delta$. Then, clearly

$$P\left(\sum_{i=0}^{n-1} X_i^{+2} < \delta n\right) \leqslant P\left(\sum_{i=0}^{n-1} \tilde{X}_i^{+2} < \delta_1 n\right) \leqslant P\left(\sum_{i=0}^{n-1} Y_i < \delta' n\right),$$

where $Y_i = K^{-1}\tilde{X}_i^{+2}$ and $\delta' = K^{-1}\delta_1$. Then $0 \leqslant Y_i \leqslant 1$ for $i \geqslant 0$. Furthermore, by the assumptions made in Section 1 and results from Doukhan (1994, see (2.4.1) and (2.4.3) on pp. 88 and 89) we have that condition (2.8) of Lemma 2.2 is satisfied by $\{K^{-1}\tilde{X}_i^{+2}\}$. Hence, the required result in (2.5) follows from (2.9) of Lemma 2.2. The result for $\{X_{i-1}^{-2}\}$ in (2.5) follows similarly.

As for (2.6), first note from (1.7), (1.3) and (1.4) that

$$\begin{aligned} Q_l &= \sum_{i=1}^{l} X_{i-1}^{+2}(\hat{\theta}_{1,l}-\theta_1)^2 + \sum_{i=1}^{l} X_{i-1}^{-2}(\hat{\theta}_{2,l}-\theta_2)^2 \\ &= Q_l^{(1)} + Q_l^{(2)}. \end{aligned} \tag{2.15}$$

We will first consider the term $Q_l^{(1)}$ in (2.15). For a $\tilde{\delta} > 0$ (to be chosen later) write

$$\begin{aligned} Q_l^{(1)} &= Q_l^{(1)} I_{\{\sum_{i=1}^{l} X_{i-1}^{+2} < \tilde{\delta} l\}} + Q_l^{(1)} I_{\{\sum_{i=1}^{l} X_{i-1}^{+2} \geqslant \tilde{\delta} l\}} \\ &\leqslant \left(\sum_{i=1}^{l} \varepsilon_i^2\right) I_{\{\sum_{i=1}^{l} X_{i-1}^{+2} < \tilde{\delta} l\}} + \tilde{\delta}^{-1} \left(\sum_{i=1}^{l} X_{i-1}^{+}\varepsilon_i \Big/ \sqrt{l}\right)^2, \end{aligned}$$

where we used (2.2) in the first term. By (2.3) we have that

$$\sup_{n\geqslant 1} E \max_{[un]\leqslant l\leqslant [vn]} \left(\sum_{i=1}^{l} X_{i-1}^{+}\varepsilon_i \Big/ \sqrt{l}\right)^{2s} < \infty. \tag{2.16}$$

By the Cauchy–Schwarz inequality and the Minkowski inequality

$$\begin{aligned} & E \max_{[un]\leqslant l\leqslant [vn]} \left(\sum_{i=1}^{l} \varepsilon_i^2\right)^{s} I_{\{\sum_{i=1}^{l} X_{i-1}^{+2} < \tilde{\delta} l\}} \\ &\leqslant E \left(\sum_{i=1}^{[vn]} \varepsilon_i^2\right)^{s} I_{\{\sum_{i=1}^{[un]} X_{i-1}^{+2} < \tilde{\delta}[vn]\}} \\ &\leqslant \mathrm{O}(n^s) \|\varepsilon_1^2\|_{2s}^{s} P^{1/2} \left\{\sum_{i=1}^{[un]} X_{i-1}^{+2} < \tilde{\delta} B_0 [un]\right\} \\ &= \mathrm{o}(1) \quad \text{as } n \to \infty, \end{aligned} \tag{2.17}$$

where we let $B_0 = [vn]/[un] (=\mathrm{O}(1))$, choose $\tilde{\delta}$ such that $\tilde{\delta} B_0 \in (0, EX_0^{+2})$ and used (2.5). Hence it follows from (2.16) and (2.17) that $\sup_{n\geqslant 1} E \max_{[un]\leqslant l\leqslant [vn]} Q_l^{(1)s} < \infty$. Similarly, it can be shown that $\sup_{n\geqslant 1} E \max_{[un]\leqslant l\leqslant [vn]} Q_l^{(2)s} < \infty$ for $Q_l^{(2)}$ defined in (2.15). Hence assertion (2.6) and the proposition *are established.* □

Lemma 2.3. *If* $E|\varepsilon_1|^{2s} < \infty$ *for* $s \geqslant 1$ *and* $n_A = \mathrm{O}(A^{1/2})$, *then for* T_A *defined in* (1.12) *and* $n_0(A)$ *defined in* (1.9)

$$\{|T_A/n_0(A)|^s; A \geqslant 1\} \text{ is u.i.} \tag{2.18}$$

Proof. Define

$$\tilde{T}_A = \inf\left\{n \geqslant n_A \colon n \geqslant (2A)^{1/2} \left(n^{-1} \sum_{i=1}^{n} \varepsilon_i^2\right)\right\}.$$

Then by Lemma 2 of Chow and Yu (1981), $\{|\tilde{T}_A/n_0(A)|^s; A \geqslant 1\}$ is u.i. However, by (2.2) we have that $T_A \leqslant \tilde{T}_A$ and hence the required result *is obtained.* □

The next lemma gives the rate of convergence of the upper- and lower-tail probability of T_A. The proof of it is omitted as it is exactly same as that of Lemma 7 of Lee (1994) [also, see Lemma 4 of Sriram (1988)].

Lemma 2.4. *If $E|\varepsilon_1|^{4s} < \infty$ for $s \geqslant 1$ and $n_A \geqslant A^{1/2(1+\eta)}$ for some $\eta > 0$. Then for $0 < \phi < 1$*

$$P\{T_A < (1-\phi)(2A)^{1/2}\sigma\} = \mathrm{O}(A^{-(s-1)/2(1+\eta)}) \tag{2.19}$$

and

$$P\{T_A > [(1+\phi)(2A)^{1/2}\sigma] + 1\} = \mathrm{o}(A^{-(s-1)/2}) \tag{2.20}$$

as $A \to \infty$.

Proof of Theorem 1. Since $\hat{\sigma}_n^2 \to \sigma^2$ a.s. as $n \to \infty$ (see Petruccelli and Woolford, 1984, for instance) the result in (1.13) follows from the definition of T_A in (1.12). This together with Lemma 2.3 yields (1.14).

For (1.15), write using the notations in (1.7),(1.5) and (1.10) that

$$R_A/R_{n_0} \approx \{AET_A^{-1}Q_{T_A} + ET_A\}/[2(2A)^{1/2}\sigma].$$

By (1.14) it suffices to show that

$$(2A)^{1/2}\sigma ET_A^{-1}Q_{T_A} \to 2\sigma^2. \tag{2.21}$$

Let $C = \{[(1-\phi)n_0(A)] \leqslant T_A \leqslant [(1+\phi)n_0(A)] + 1\}$ for $\phi \in (0,1)$. By the Cauchy–Schwarz inequality, (2.6), and (2.19) and letting $n_1(A) = [(1-\phi)n_0(A)]$ we have

$$\begin{aligned}(2A)^{1/2}\sigma ET_A^{-1}Q_{T_A}I_{\{T<n_1(A)\}} &\leqslant (2A)^{1/2}\sigma||T_A^{-1}Q_{T_A}I_{\{n_A\leqslant T\leqslant n_1(A)\}}||_2 \\ &\quad\times P^{1/2}(T_A \leqslant n_1(A)) \\ &\leqslant (2A)^{1/2}\sigma \sup_n ||Q_n||_2 \left(\sum_{n=n_A}^{\infty} n^{-2}\right)^{1/2} P^{1/2}(T_A \leqslant n_1(A)) \\ &= A^{1/2}\mathrm{O}(n_A^{1/2})\mathrm{O}(A^{-(s-1)/4(1+\eta)}) \\ &\to 0 \quad \text{as } A \to \infty. \end{aligned} \tag{2.22}$$

Similar arguments using (2.20) yields $(2A)^{1/2}\sigma ET_A^{-1}Q_{T_A}I_{\{T>n_2(A)\}} \to 0$ as $A \to \infty$ where $n_2(A) = [(1+\phi)n_0(A)] + 1$. Also, by (2.4), the fact that $n^{-1}\sum_{i=1}^n X_{i-1}^{\pm 2} \to EX_0^2$ a.s., and repeated application of Lemma 1.4 of Woodroofe (1982) it follows that $\{Q_n; n \geqslant 1\}$ is u.c.i.p. Therefore, from (1.6), (1.13) and the Anscombe's theorem it follows that

$$(2A)^{1/2}\sigma T_A^{-1}Q_{T_A}I_C \xrightarrow{\mathrm{D}} \sigma^2\chi_2^2 \quad \text{as } n \to \infty. \tag{2.23}$$

Moreover, by (2.6) we have for any $\beta > 1$

$$\begin{aligned} E[(2A)^{1/2}\sigma T_A^{-1}Q_{T_A}I_C]^\beta &\leqslant [n_1(A)]^{-\beta}(2A)^{\beta/2}\sigma^\beta E \max_{n_1(A)\leqslant n\leqslant n_2(A)} Q_n^\beta \\ &= \mathrm{O}(1). \end{aligned}$$

Therefore, $\{(2A)^{1/2}\sigma T_A^{-1}Q_{T_A}I_C; A \geqslant 1\}$ is u.i. and hence

$$(2A)^{1/2}\sigma ET_A^{-1}Q_{T_A}I_C \to 2\sigma^2 \quad \text{as } A \to \infty. \tag{2.24}$$

The result in (2.21) now follows from (2.22) to (2.24). Hence the theorem *is established.* □

3. Asymptotic distribution and second-order expansion

In this section we prove Theorem 2 stated in Section 1 using results from nonlinear renewal theory developed by Lai and Siegmund (1977,1979) and Hagwood and Woodroofe (1982). The reader is referred to Woodroofe (1982) for thorough exposition of the theory.

First, rewrite the stopping rule defined in (1.12) as

$$\begin{aligned} T_A &= \inf\{n \geqslant n_A \colon n(\sigma/\hat{\sigma}_n) \geqslant (2A)^{1/2}\sigma\} \\ &= \inf\{n \geqslant n_A \colon S_n + \xi_n \geqslant (2A)^{1/2}\sigma\}, \end{aligned} \tag{3.1}$$

where

$$S_n = \sum_{i=1}^{n} [1 - (1/2)(\varepsilon_i^2/\sigma^2 - 1)]$$

and

$$\xi_n = (1/2)Q_n/\sigma^2 + (3/8)\lambda_n^{-5/2} n(\hat{\sigma}_n^2/\sigma^2 - 1)^2, \tag{3.2}$$

where λ_n is a random variable such that $|\lambda_n - 1| < |\hat{\sigma}_n^2/\sigma^2 - 1|$. The last equality in (3.1) is obtained by expanding $(\sigma^2/\hat{\sigma}_n^2)^{1/2}$ around 1 using Taylor's theorem and substituting the identity for $\hat{\sigma}_n^2$ in (2.1). Clearly, S_n defined in (3.2) is a sum of i.i.d. random variables with mean 1 and the sequence $\{\xi_n, n \geqslant 1\}$ can be shown to be slowly changing (see, Woodroofe, 1982, for a definition). In fact, using arguments similar to those in Section 3 of Sriram (1988) it is possible to show that the two terms in ξ_n denoted by

$$L_n^1 = (1/2)Q_n/\sigma^2 \quad \text{and} \quad L_n^2 = (3/8)\lambda_n^{-5/2} n(\hat{\sigma}_n^2/\sigma^2 - 1)^2 \tag{3.3}$$

slowly change sequences as well. With these observations we proceed to prove Theorem 2.

Proof of Theorem 2. In view of the discussions above the asymptotic normality of the standardized stopping rule in (1.16) would follow from Lemma 4.2 of Woodroofe (1982) provided we show that $\xi_n/\sqrt{n} \to 0$ in probability. From (3.2), (3.3), (2.1), (1.6) and the strong consistency of $\hat{\sigma}_n^2$ defined in (1.11) we have that $\xi_n/\sqrt{n} \to 0$ in probability. Hence the assertion in (1.16) *is obtained.*

As for the second-order expansion for the expected value of T_A defined in (1.12) we will use a slightly general version of Theorem 4.5 of Woodroofe (1982) given in Sriram (1988); see Lemma 7. In Lemma 7 of Sriram (1988), set A_n, l_n and μ to be

$\Omega, 0$ and 1. Also let $\tilde{\xi}_n = \xi_n$ defined in (3.2) with L_n^1 and L_n^2 as defined in (3.3). Then, by (1.6)

$$L_n^1 \xrightarrow{\mathrm{D}} (\tfrac{1}{2})\chi_2^2 = L_1 \quad \text{say as } n \to \infty, \tag{3.4}$$

and by (2.1), CLT, the fact that $Q_n/\sqrt{n} \to 0$ in probability and $\lambda_n \to 1$ in probability we have that

$$L_n^2 \to (3/8)\chi_1^2 Var(\varepsilon_1^2/\sigma^2) = L_2, \quad \text{say as } n \to \infty. \tag{3.5}$$

Therefore, conditions (3.7)–(3.9) of Lemma 7 of Sriram (1988) are satisfied.

As for the verification of (3.10) of Lemma 7 of Sriram (1988), for the set C defined in the proof of Theorem 1 above, it readily follows from (2.6) of Proposition 2.1 that

$$\left\{ \max_{n_1(A) \leqslant l \leqslant n_2(A)} L_l^1 I_C; A \geqslant 1 \right\} \text{ is u.i.} \tag{3.6}$$

Also,

$$\max_{n_1(A) \leqslant l \leqslant n_2(A)} L_n^2 I_C \leqslant (3/8) \max_{n_1(A) \leqslant l \leqslant n_2(A)} \lambda_l^{-5/2} l(\hat{\sigma}_l^2/\sigma^2 - 1)^2 I_C, \tag{3.7}$$

where, by the definition of λ_l in (3.2),

$$\max_{n_1(A) \leqslant l \leqslant n_2(A)} \lambda_l^{-5/2} I_C \leqslant 1 + \max_{n_1(A) \leqslant l \leqslant n_2(A)} (\sigma^2/\hat{\sigma}_l^2)^{5/2} I_C. \tag{3.8}$$

Let $d_l = l\hat{\sigma}_l^2$ for $l \geqslant 1$ where $\hat{\sigma}_l^2$ is defined in (1.11). Then, since $\hat{\theta}_{1,n}$ and $\hat{\theta}_{2,n}$ in (1.3) and (1.4) are least-squares estimators, we have that $d_l \leqslant d_{l+1}$ for $l \geqslant 1$. Therefore,

$$\begin{aligned} \max_{n_1(A) \leqslant l \leqslant n_2(A)} (\sigma^2/\hat{\sigma}_l^2)^{5/2} I_C &\leqslant \mathrm{O}(1)[(n_1(A) - 1)\sigma^2/d_{n_1(A)-1}]^{5/2} I_C \\ &= \mathrm{O}(1), \end{aligned} \tag{3.9}$$

since the definition of T_A implies that

$$C \subset \{[n_1(A) - 1] < (2A)^{1/2} \hat{\sigma}_{n_1(A)-1}\} = G \tag{3.10}$$

and on G

$$\begin{aligned} \{[n_1(A) - 1]\sigma^2/d_{n_1(A)-1}\}^{5/2} I_G &\leqslant \{(2A)^{1/2}/[n_1(A) - 1]\}^5 I_G \\ &= \mathrm{O}(1), \end{aligned} \tag{3.11}$$

where the fact used is that $n_1(A) = \mathrm{O}(A^{1/2})$. From (3.8) to (3.11) we have that

$$\max_{n_1(A) \leqslant l \leqslant n_2(A)} L_l^2 I_C \leqslant \mathrm{O}(1) \max_{n_1(A) \leqslant l \leqslant n_2(A)} l(\hat{\sigma}_l^2/\sigma^2 - 1)^2 I_C. \tag{3.12}$$

From Doob's maximal inequality (see Chow and Teicher, 1978, Corollary 10.3.2) it can be shown that

$$\left\{ \max_{n_1(A) \leqslant l \leqslant n_2(A)} l^{-1} \left[\sum_{i=1}^{l} (\varepsilon_i^2/\sigma^2 - 1) \right]^2 ; A \geqslant 1 \right\} \text{ is u.i.} \tag{3.13}$$

Also,

$$\begin{aligned} E \max_{n_1(A) \leqslant l \leqslant n_2(A)} l^{-1} Q_l^2 &\leqslant [n_1(A)]^{-1} E \max_{n_1(A) \leqslant l \leqslant n_2(A)} Q_l^2 \\ &\to 0, \end{aligned} \tag{3.14}$$

by (2.6). Hence, it follows from (3.13), (3.14), (2.1) and (3.12) that

$$\left\{ \max_{n_1(A)\leqslant l\leqslant n_2(A)} L_l^2 I_C; A\geqslant 1 \right\} \text{ is u.i.} \tag{3.15}$$

The verification of condition (3.10) of Sriram (1988) now follows from (3.6) and (3.15). Condition (3.11) of Sriram (1988) is trivially satisfied and condition (3.12) follows from (2.19). Therefore, by the conclusion of Lemma 7 of Sriram (1988), (3.4) and (3.5) we have that

$$ET_A = (2A)^{1/2}\sigma + \rho - 1 - (3/8)Var(\varepsilon_1^2/\sigma^2) + o(1) \tag{3.16}$$

as $A \to \infty$, where $\rho = ES_{\tau_0}^2/2ES_{\tau_0}$ for S_n defined in (3.2) and $\tau_0 = \inf\{n \geqslant n_A\colon S_n \geqslant 0\}$. Hence the conclusion in (1.17) and the theorem *are established.* □

Acknowledgements

We would like to thank the two referees for their constructive comments and suggestions which improved the presentation of the paper.

References

Chan, K.S., Petruccelli, J.D., Tong, H., Woolford, S.W., 1985. A multiple threshold AR(1) model. J. Appl. Probab. 22, 267–279.

Chow, Y.S., Teicher, H., 1978. Probability theory: Independence, Interchangeability, Martingales. Springer, New York.

Chow, Y.S., Yu, K.F., 1981. The performance of a sequential procedure for the estimation of the mean. Ann. Statist. 9, 184–189.

Doukhan, P., 1994. Mixing Properties and Examples. Lecture notes in Statistics. Springer, New York.

Fakhre-Zakeri, I., Lee, S., 1992. Sequential estimation of the mean of a linear process. Sequential Anal. 11, 181–197.

Fakhre-Zakeri, I., Lee, S., 1993. Sequential estimation of the mean vector of a multivariate linear process. J. Multivariate Anal. 47, 196–209.

Freedman, D., 1973. Another note on the Borel–Cantelli lemma and the strong law, with the Poisson approximation as a by product. Ann. Probab. 1, 910–925.

Ghosh, M., Mukhopadhyay, N., Sen, P.K., 1997. Sequential Estimation. Wiley, New York.

Hagwood, C., Woodroofe, M., 1982. On the expansion for expected sample size in non-linear renewal theory. Ann. Probab. 10, 844–848.

Lai, T.L., Siegmund, D., 1977. A non-linear renewal theory with applications to sequential analysis. I. Ann. Statist. 5, 946–954.

Lai, T.L., Siegmund, D., 1979. A non-linear renewal theory with applications to sequential analysis. II. Ann. Statist. 7, 60–76.

Lee, S., 1994. Sequential estimation for the parameters of a stationary autoregressive model. Sequential Anal. 13, 301–317.

Lee, S., 1996. Sequential estimation for the autocorrelations of linear processes. Ann. Statist. 24, 2233–2249.

Martinsek, A.T., 1983. Second order approximation to the risk of a sequential procedure. Ann. Statist. 11, 827–836.

Petruccelli, J.D., Woolford, S.W., 1984. A threshold AR(1) model. J. Appl. Probab. 21, 207–286.

Robbins, H.E., 1959. Sequential estimation of the mean of a normal population. In: U. Grenander (Ed.), Probability and Statistics. Wiley, New York, pp. 235–245.

Sriram, T.N., 1987. Sequential estimation of the mean of a first order stationary autoregressive process. Ann. Statist. 15, 1079–1090.

Sriram, T.N., 1988. Sequential estimation of the autoregressive parameter in a first order autoregressive process. Sequential Anal. 7, 53–74.

Sriram, T.N., 1998. Fixed size confidence regions for parameters of threshold AR(1) models. Submitted for publication.

Sriram, T.N., Basawa, I.V., Huggins, R.M., 1991. Sequential estimation for branching processes with immigration. Ann. Statist. 19, 2232–2243.

Tong, H., 1978a. On a threshold model. In: C.H. Chen (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, Amsterdam.

Tong, H., 1978b. Threshold autoregression, limit cycles and cyclical data. Tech. Rep. No. 101, Dept. Math. Stat., UMIST.

Tong, H., 1983. Threshold Models in Non-linear Time Series Analysis, Lecture Notes in Statistics, Vol. 21, Springer, Berlin.

Tong, H., 1990. Nonlinear Time Series. Oxford University Press, Oxford.

Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data (with Discussion). J. R. Stat. Soc. B 42, 245–292.

Woodroofe, M., 1982. Nonlinear Renewal Theory in Sequential Analysis. SIAM, Philadelphia.